

G6PD variant distribution in sub-Saharan Africa and potential risks of using chloroquine/hydroxychloroquine based treatments for COVID-19

Jorge da Rocha^{1,2,‡,*}, Houcemeddine Othman^{1,‡}, Caroline T. Tiemessen³, Gerrit Botha^{4,‡}, Michèle Ramsay^{1,2,‡}, Collen Masimirembwa¹, Clement Adebamowo^{5,6,‡}, Ananyo Choudhury^{1,‡}, Jean-Tristan Brandenburg^{1,‡}, Mogomotsi Matshaba^{8,9,‡}, Gustave Simo^{7,‡}, Francisco-Javier Gamo¹⁰ and Scott Hazelhurst^{11,1,‡} as members of the H3Africa Consortium[‡].

1 Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

2 Division of Human Genetics, National Health Laboratory Service, and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

3 Centre for HIV and STIs, National Institute for Communicable Diseases, National Health Laboratory Services and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg South Africa.

4 Computational Biology Division and H3ABioNet, Department of Integrative Biomedical Sciences, University of Cape Town, South Africa.

5 Institute for Human Virology, Abuja, Nigeria.

6 Institute of Human Virology and Greenebaum Comprehensive Cancer Center, University of Maryland School of Medicine, Baltimore, MD

7 Molecular Parasitology and Entomology Unit, Department of Biochemistry, Faculty of Science, University of Dschang, Dschang, Cameroon.

8 Botswana-Baylor Children's Clinical Center of Excellence, Gaborone, Botswana.

9 Baylor College of Medicine, Houston, United States.

10 Global Health, GlaxoSmithKline R&D, Madrid, Spain.

11 School of Electrical & Information Engineering, University of the Witwatersrand, Johannesburg, South Africa.

‡Members of the Human Heredity and Health in Africa Consortium.

* jorge.darocho@students.wits.ac.za

Abstract

Chloroquine/hydroxychloroquine have been proposed as potential treatments for COVID-19. These drugs have warning labels for use in individuals with glucose-6-phosphate dehydrogenase (G6PD) deficiency. Analysis of whole-genome sequence data of 458 individuals from sub-Saharan Africa showed significant *G6PD* variation across the continent. We identified nine variants, of which four are potentially deleterious to G6PD function, and one (rs1050828) that is known to cause G6PD deficiency. We supplemented data for the rs1050828 variant with genotype array data from over 11,000 Africans. Although this variant is common in

Africans overall, large allele frequency differences exist between sub-populations. African sub-populations in the same country can show significant differences in allele frequency (e.g. 16.0% in Tsonga vs 0.8% in Xhosa, both in South Africa, $p = 2.4 \times 10^{-3}$). The high prevalence of variants in the *G6PD* gene found in this analysis suggests that it may be a significant interaction factor in clinical trials of chloroquine and hydrochloroquine for treatment of COVID-19 in Africans.

1 Introduction

Chloroquine and hydroxychloroquine (CQ/HCQ) are currently undergoing clinical trials as treatments for Coronavirus disease 2019 (COVID-19) which is caused by the severe acute respiratory syndrome coronavirus 2 – SARS-Cov-2 [32]. CQ/HCQ and other aminoquinolines have pharmacogenomic associations with the glucose-6-phosphate dehydrogenase (*G6PD*) gene [23]. Aminoquinolines are suspected to exert their antimalarial effect by increasing oxidative stress via production of haem-based reactive oxygen species [16]. The *G6PD* enzyme is responsible for the production of nicotinamide adenine dinucleotide phosphate (NADPH) which is required in the glutathione mediated detoxification of reactive oxygen species [21]. In the case of inactive/deficient *G6PD*, the NADPH supply may not be sufficient to neutralize the reactive oxygen species induced by CQ/HCQ and other drugs with similar mechanisms of action.

G6PD deficiency is common globally, particularly in African populations (14% of males) [27]. Individuals with the deficiency are at risk for haemolytic anaemia which can be triggered by infections, certain foods or medications. *G6PD* deficiency is an X-linked disorder. It mostly occurs in males who are hemizygous for deleterious variants of the *G6PD* gene and in females with homozygous deleterious variants. Symptoms have also been observed in females with heterozygous combinations due to X-inactivation effects [40]. Three common haplotype arrangements have been defined for the gene, the B (wild type), A and A⁻ (deficiency). The *G6PD* A 'haplotype' is denoted by the rs1050829 variant (c.376A>G). The rs1050829 C variant is not linked to decreased *G6PD* activity and occurs in 10 - 30% of sub-Saharan Africans

[38, 19].

The A- haplotype (which is associated with G6PD enzyme deficiency) is formed by a combination of two variants, one of which is a deleterious variant, while the other is the rs1050829 C variant. A commonly observed combination is that of rs1050828 T (c.202G>A), which occurs in 10% of sub-saharan Africans, and rs1050829 [23]. The *G6PD* A- haplotype causes between 10-60% reduction in G6PD enzyme activity. The A- *G6PD* haplotype is classified as a World Health Organisation (WHO) Class III variant [1, 23]. Strong linkage disequilibrium exists between the rs1050828 T and rs1050829 C variants [39]. As the rs1050829 C allele is more common, it is likely that that rs1050828 T emerged after rs1050829 C, and then increased in frequency due to positive selection in Africans [33]. As rs1050829 C has no effect on G6PD deficiency, it is reasonable to report G6PD deficiency based on rs1050828 genotype combinations alone.

The FDA has issued warnings on the use of CQ and HCQ in G6PD deficient individuals due to high risk of haemolytic anaemia, although these are not contraindicated [36, 37]. Acute haemolytic effects following HQC treatment for COVID-19 have been reported in a single male case [8] who is suspected of carrying the *G6PD* Mediterranean variant. Whereas CQ is not known to induce severe haemolytic effects when used as an antimalarial in G6PD deficient individuals, in contrast to primaquine [34] or chlorproguanil [9], the risk of its therapeutic use in G6PD deficient, COVID-19 patients is unknown.

In this paper, we evaluate the prevalence of variants in *G6PD* gene in individuals of African ancestry. We suggest that variations in the *G6PD* gene could significantly affect risk of adverse effects of CQ/HCQ, and recommend that this should be evaluated in clinical trials of CQ/HCQ treatment for COVID-19. We also report the prevalence of a key *G6PD* variant, rs1050828, in 11,030 Africans from four countries in west, east and southern Africa, and show that not only is the variant allele common in Africa overall, but that there are very large differences between different groups, even between those who reside in close proximity.

2 Methods

The dataset used was assembled as a collaborative project of the Human Heredity and Health in Africa (H3A) Consortium. The high coverage African ADME Dataset (HAAD) was sourced from H3A, other African collaborations and the Simon Foundation's Genome Diversity Project [18]. HAAD consists of high-coverage sequences from 458 Sub-Saharan African individuals from 15 countries, with 8 of these countries contributing data from more than 25 individuals (Nigeria, Ghana, Burkina Faso, Cameroon, Benin, Botswana, Zambia, and South Africa) (Figure 1 A). HAAD BAMs were aligned to GRCh37 with bwa-mem v0.7.10- v0.7.17 [14]. Variants were called with Haplotype-Caller in gVCF mode using GATK v.4.0.8.1 HAAD gVCFs (along with gVCFs produced with African 1000 Genomes Project data (KGA) [2]) were combined with GATK's CombineGVCF (v.4.0.8.1), and jointly called with GenotypeGVCFs (v4.1.3.0) and followed GATK's best practice guidelines. VQSR was used to select high quality sites with PASS ratings. All related workflows for data preparation can be found at <https://github.com/h3abionet/recalling>. The *G6PD* canonical gene region (chrX:153759606-153775469) was extracted with bcftools v1.9, and variants were annotated (e.g. as missense, intronic etc.) with variant effect predictor (VEP) v92.0 [24] and SNPeff v4.3t [6]. Coding variants were selected for analysis if they meet a *QUAL* > 50 quality score. Functional annotation for these variants was performed using dbNSFP [15] to retrieve scores for five predictive toolsets (LRT, MutationAssessor, PROVEAN, VEST3 and CADD), which were then used for prediction based on a pharmacogene optimised model [43]. These form part of the "g_miner" workflow which is available at: https://github.com/hothman/PGx-Tools/tree/master/workflows/g_miner. PLINK [4] was used to call allele frequencies in every country in the HAAD dataset. Statistical analyses were conducted using R v3.63 [28]. The test for equal or given proportions was used to calculate allele frequency confidence intervals (CI) at the 95% significance level. Fisher's exact test was used to assess significant differences in allele frequency between two populations, at the 5% significance threshold.

The impact of the variant on the protein structure was assessed by DynaMut [31] using the crystal structure of Canton G6PD modified to the wild type form [12]. Stability related metrics

calculated by Dynamut include the change in Gibbs free energy ($\Delta\Delta G$).

The A- haplotype in this study is defined by the presence of the rs1050828 T allele (alleles assessed in forward orientation, this corresponds to the c.202G>A nomenclature for cDNA NM_001042351.2). The presence of rs1050829 C (corresponding to c.376A>G) is assumed due to strong linkage disequilibrium with rs1050828 T [39]. The rs1050828 genotypes were previously generated by the AWI-Gen Project [30] on 11,062 sub-Saharan Africans from 6 sites in Ghana, Burkina Faso, Kenya and South Africa using the H3A Custom Genotyping Array (<https://www.h3abionet.org/h3africa-chip>). PLINK [4] was used to remove any individuals with more than 1% missingness or genotypes which conflicted with declared sex. This left 11,030 individuals in all. The cluster plots of the genotype calls (male, female, all) are consistent with a well-genotyped single nucleotide polymorphism on the X chromosome. Minor allele frequency (MAF) was computed for the overall AWI-Gen data set as well as various sub-groups, as determined by self-identified ethnicity. The data was analysed using a custom Python script using the pandas-plink library (<https://github.com/limix/pandas-plink>).

3 Results

3.1 Variation from high coverage data

Nine coding (missense) variants were identified in the WGS, of which seven have been previously described, and have rsIDs allocated on dbSNP (nomenclature referred to by rsID throughout, as defined in Table 1). No loss of function type variants were detected.

Figure 1B shows the distribution of the *G6PD* missense variants across African populations, along with comparisons to their overall frequency in the African data from the Genome Aggregation Database (gnomAD) [13] and the 1000 Genomes Project (KGP) [2]. Of these nine variants, seven have at least one prediction score from the pharmacogene model indicating deleterious impact. The variants rs1050829 (chrX:g.153763492T>C), and chrX:g.153775028C>T had no predictive score reaching model cutoff criteria.

G6PD deficiency (A-), as defined by the rs1050828 T allele, is widely distributed across the African continent (Figure 1 A). There are notable differences in frequency across different

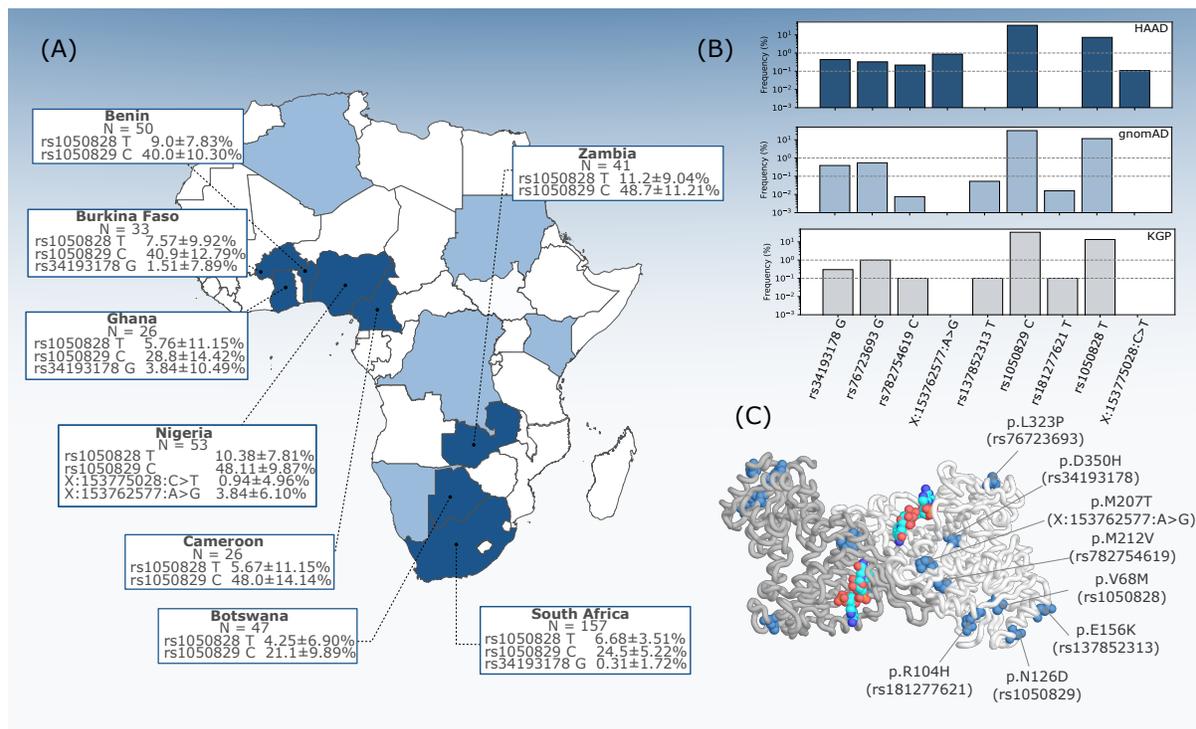


Figure 1: *G6PD* missense variant distribution across African populations. (A) *G6PD* allele frequencies in populations from HAAD countries. Confidence intervals for allele frequencies based on the equal or given proportions test the 95% significance level. (B) Allele frequencies of missense variants in HAAD, and African superpopulation groups from gnomAD and the KGP. (C) Structural representation of the *G6PD* homodimer with missense residues highlighted in blue color on both chains with bound NADP (NADP shown in red-turquoise-blue).

groups, and we note that frequency does not necessarily correlate with the geographic location of the country. The highest frequency observed in HAAD populations was in Zambians ($11.2 \pm 9.04\%$, 95% CI), whereas the lowest was in Botswana ($4.25 \pm 6.90\%$, 95% CI) their geographic neighbours. At the low sample number, this difference is not significant ($p = 0.09115$), and the confidence interval is large, however, we note that allele frequency is not uniform across these and other HAAD African groups.

Two other missense variants in HAAD, rs76723693 (chrX:g.153761240A>G) and rs34193178 (chrX:g.153761160C>G), have robust predictions as being functionally deleterious (all model tools in consensus). rs76723693 is defined as part of the A- haplotype and although very rare in Africans, it has been characterised in other global populations [23]. The rs34193178 variant is rare overall, but is found at a $1.5 \pm 7.89\%$ in HAAD Burkina Faso and $3.84 \pm 10.49\%$ in HAAD

Ghana populations (95% CI). It is also present in HAAD South Africans at $0.3 \pm 1.72\%$ (95% CI).

The two variants chrX:g.153762577A>G and chrX:g.153775028C>T were not found in dbSNP151, and have not been reported in the gnomAD and the KGP databases (Figure 1B). The chrX:g.153762577A>G variant is only found in the HAAD Nigerian population, at $3.8 \pm 6.10\%$ (95% CI), but not seen in the KGA Nigerian Esan or Yoruba populations. The chrX:g.153775028C>T is a singleton variant, and is only present in pre-protein structures and is thus not displayed in Figure 1C.

Four of the variants (rs1050828, rs137852313, rs1050829, rs181277621) are located in the co-enzyme domain; rs782754619 and chrX:g.153762577A>G belong to the $\alpha + \beta$ domain of G6PD buried in the protein core; and rs76723693 and rs34193178 are exposed to the solvent. All the corresponding amino acid residues (with the exception of rs137852313) are either densely packed against other residues within the structure of G6PD, or they establish polar contacts that appear to stabilize local conformations of nearby segments. Structural predictions (as based on $\Delta\Delta G$ (kcal/mol) - Table 1) of protein variant effect show a destabilizing effect for rs76723693, rs782754619, chrX:g.153762577A>G, rs181277621, and rs1050828. The rs34193178, rs137852313 and rs1050829 variants showed a stabilizing effect.

Table 1: G6PD missense variants detected within HAAD and KGA population datasets and their relative stability effect ($\Delta\Delta G$) for G6PD protein.

Variant ID	Nucleotide	Amino Acid	$\Delta\Delta G$ (kcal/mol)
rs34193178	1048G>A	D350H	0.374
rs76723693	968T>C	L323P	-0.872
rs782754619	634A>G	M212V	-1.147
rs181277621	311G>A	R104H	-1.401
chrX:g.153762577A>G	T620C	M207T	-1.145
chrX:g.153775028C>T	G58A	N/A	N/A
rs137852313	466G>A	E156K	0.539
rs1050829	376A>G	N126D	0.887
rs1050828	202G>A	V68M	-1.347

Nucleotide positions based on cDNA NM_001042351.2 cDNA (chrX:g.153775028C>T G58A is indicated by NM_000402.4). Amino acid positions based on the protein sequence of from 6E07 PDB structure.

3.2 High variability of rs1050828 allele frequency in Africa

Previous studies have shown that the MAF of rs1050828 is relatively high in African populations [20, 7]. We further show that it is also extremely variable – even within the same geographical region. Table 2 shows the minor allele frequency in the AWI-Gen study (over 11,030 participants). A full per-group analysis is not possible in this rapid communication. However, we show the MAF in selected groups with at least 180 individuals.

The rs1050828 T variant frequency in 11,030 individuals was 12.6%, close to the values reported in gnomAD (11.6%) and KGP (13.5%). Overall, 11.9% of African males carried the variant on their X chromosome while 2.2% of females were homozygous for the T allele. A moderate *G6PD* deficiency (10-60% residual enzyme activity) (WHO Class III) is likely to be present in individuals with such genotypes [1]. Although rare, the deficiency can present in heterozygous females depending on X-inactivation effects [40]. There was a significant difference in the variant frequency among self-identified ethnic groups in South Africa and Kenya. The frequency among the Tsonga was 16.0% which is substantially different from 0.8% found in the Xhosa ($p = 2.4 \times 10^{-3}$) and 5.5% in the BaPedi-Tswana-Sotho ($p = 2.2 \times 10^{-16}$) ethnolinguistic groups. The T variant appears at markedly different frequency in different Kenyan groups – 6.8% in the Kikuyu and 11% in the combined Luhya-Luo-Kamba groups ($p = 4.3 \times 10^{-3}$).

We tested for deviation from Hardy-Weinberg equilibrium (HWE) in the females in each of the groups except for the Xhosa (in which there was only 1 person who had the variant allele) and no significant deviation could be shown (lowest p -value was 0.38). In women overall the expected proportion of women who are heterozygous under assumption of HWE is 0.224 while the observed proportion was 0.213. While this is highly significant ($p = 2.7 \times 10^{-4}$), this deviation is not unexpected given the very significant differences in MAF between groups.

Table 2: Minor allele frequency (MAF) of rs1050828 (T) in selected groups from the AWI-Gen study genotype data. Het: Proportion of females heterozygous (%), Hom: Proportion of females homozygous for the alternate allele (%). Note that 100 of the HAAD SA individuals are included in this genotyping study (< 2% of the samples)

Group	All		Males	Females			
	N	MAF	MAF	N	MAF	Het	Hom
All Genotyped Samples	11,030	12.6	11.9	6033	12.8	21.3	2.2
South Africa							
Tsonga	2,132	16.0	14.1	1209	16.7	26.9	3.2
BaPedi-Tswana-Sotho	1,849	5.5	5.1	1233	5.6	10.5	0.4
Xhosa	180	0.8	0.9	68	0.7	1.5	0.0
Ghana and Burkina Faso							
Mampruga-Mossi-Gouronsi-Kassena-Nankana	3,723	18.6	17.6	1,888	19.1	30.7	3.7
Kenya							
Luhya-Luo-Kamba	978	11.0	8.6	454	12.3	20.7	2.0
Kikuyu	655	6.8	6.3	434	6.9	12.4	0.7

Discussion

Current clinical studies that have used CQ/HCQ for treatment of COVID-19 have not explicitly taken into account the potential risks posed by G6PD deficiency [3, 5, 10]. G6PD deficiency is known to be common in Africans. In the present study, we assessed *G6PD* gene variation in African populations, and noted the high prevalence of a common deleterious allele – rs1050828. Although common, there are large differences in frequency for this variant, even between populations that are geographic neighbours. The differences may be explained by selective pressures in regions where malaria is/was common [11], as G6PD deficiency may convey resistance to malaria [22]. Such differences have been previously reported to occur even within countries, as in Botswana, where a decreasing trend in frequency of this variant occurs from the north-west to south-east [26]. Another study in South Africans (n=181) from the Mpumalanga province reported the allele frequency of A- to be 14% [29], which is similar to findings from AWI-Gen genotype data, where a MAF of 16% was found in Tsonga-speaking individuals. However, other SA groups, such as the Xhosa in particular (MAF 0.8%) had much lower frequencies. This highlights the limitation of reporting allele frequency by country rather than ethnolinguistic groups. African populations undergoing COVID-19 CQ/HCQ treatment

trials may not have the same relative frequency of this allele as others. Thus we urge that models for *G6PD* related effects based on a single proxy African population are not directly transferable to other Africans, even those close geographically.

We observed other potentially deleterious variants in the African populations we studied. For instance, rs34193178 and chrX:g.153762577A>G, which were not found across all populations, do not have well characterised effects, and are unlikely to be included in assays to type well known *G6PD* variants. If these have a functional impact on G6DPD, they may add complexities to studies assessing the presence of rs1050828 and rs1050829 alone. The chrX:g.153762577A>G variant in particular has structural evidence for deleterious functional impact.

Although use of either CQ/HCQ is not new in African populations, the dosage and duration of CQ/HCQ treatment for COVID-19 may lead to higher prevalence of adverse effects related to G6PD deficiency. Acute haemolysis has been observed in a G6PD deficient male suspected of carrying the *G6PD* Mediterranean variant (rs5030868) who was treated with lopinavir and HCQ [8]. Although the Mediterranean variant is known to induce a greater sensitivity to pro-oxidant drugs than the A- variant [43], the observation of haemolytic effects highlights the relevance of assessing the impact of G6PD deficiency on CQ/HCQ that is repurposed for COVID-19 treatment. Non-haemolytic adverse effects have also been noted in other recent trials. A CQ trial in Brazilians noted severe adverse reactions related to QT elongation [3], and the high doses (600 mg twice daily) may pose greater risks for individuals with G6PD deficiency. The proportion of African admixture for the patients assessed in this study was not disclosed. A recent trial in U.S. Veterans showed increased risk of mortality in patients treated with HQC [17]. In addition, it is currently unknown how G6PD deficiency may affect COVID-19 disease progression. G6PD deficient cells have been found to be more vulnerable to human *alphacoronavirus* 229E infection in vitro, which correlated with elevated oxidant production [41], although it is not yet known if this effect would also be seen with the novel SARS-Cov-2 virus. Monitoring of G6PD deficiency throughout COVID-19 trials and studies in Africans may therefore also reveal other factors which are not limited only to effects of drug response. Such studies could make use of a rapid enzymatic assay for G6PD deficiency in lieu

of DNA-based assays. These are available, though it should be noted that their sensitivity is lower in females [42].

As a final note, the applicability and evidence basis of CQ/HCQ as COVID-19 treatments have recently been well reviewed [35, 25] and their current implementation have been questioned [35]. These reviews conclude that there is currently insufficient evidence for the use of CQ/HCQ as COVID-19 treatments. As trials are still ongoing, we urge the consideration of G6PD deficiency related effects in African populations participating in these studies.

4 Conclusion

Given our findings of the large heterogeneity of the *G6PD* gene, variants associated with G6PD deficiency in Sub-Saharan Africa, and the possible presence of other uncharacterised deleterious variants, it is important to consider the potential impact of these variants before widespread use of CQ/HCQ as COVID-19 treatments for African populations. Distinct African ethnolinguistic groups can have vastly different frequencies of G6PD deficiency, thus clinical studies of CQ/HCQ for COVID-19 in Africans should be conducted on diverse African populations, and with monitoring for haemolysis and/or anaemia. Targeted sequencing of *G6PD* among study participants would provide important insights into the risks of adverse effects at therapeutic doses which might lead to dosage adjustment.

Acknowledgements

This paper uses the results of data and analysis from the GSK/H3A ADME project. We acknowledge the generous collaboration of the H3A AWI-Gen, CAfGEN, and TrypanoGEN groups, the support of the Cell Biology Research Lab, NICD/Wits, the Awadalla Lab at the Université de Montréal, Gustave Simo from the University of Dschang, Clement Adebamowo at the Institute for Human Virology, Abuja, Martin Simuunza from the Department of Disease Control at the University of Zambia, and Gabriel Anabawi from the University of Botswana. We sincerely thank the data providers, and study participants of these groups for their contribution. The AWI-Gen Collaborative Centre is funded by the NIH/NHGRI (Grant U54HG006938) as

part of the H3Africa Consortium. MR is a South African Research Chair in Genomics and Bioinformatics of African Populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology, and administered by National Research Foundation of South Africa (NRF). The TrypanoGEN project was funded by the Wellcome Trust, study number 099310/Z/12/Z. The Collaborative African Genetics Network (CAfGEN) is funded by the NIH/NHGRI (grant CAfGEN 1U54AI110398). The whole genome sequencing of the H3A Data was supported by a grant from the National Human Genome Research Institute, National Institutes of Health (NIH/NHGRI) U54HG003273. Cell Biology Research Lab component: This work is based on the research supported by grants awards from the Strategic Health Innovation Partnerships (SHIP) Unit of the South African Medical Research Council, a grantee of the Bill & Melinda Gates Foundation, and the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa (84177). JdR was partially funded by the SA National Research Foundation (SFH170626244782). GSK had no part in the design of this study or analysis presented here and the views and opinions expressed are not necessarily those of GSK. The authors acknowledge the financial support of GSK.

References

- [1] Glucose-6-phosphate dehydrogenase deficiency. WHO Working Group. *Bull. World Health Organ.*, 67(6):601–611, 1989.
- [2] A. Auton, L. D. Brooks, and R. M. Durbin *et al.* A global reference for human genetic variation. *Nature*, 526(7571):68–74, Oct 2015.
- [3] Mayla Gabriela Silva Borba, Fernando Fonseca Almeida Val, Vanderson Souza Sampaio, Marcia Almeida Araújo Alexandre, Gisely Cardoso Melo, Marcelo Brito, Maria Paula Gomes Mourão, José Diego Brito-Sousa, Djane Baía-da Silva, Marcus Vinitius Farias Guerra, Ludhmila Abrahão Hajjar, Rosemary Costa Pinto, Antonio Alcirley Silva Balieiro, Antônio Guilherme Fonseca Pacheco, Jr Santos, James Dean Oliveira, Felipe Gomes Naveca, Mariana Simão Xavier, André Machado Siqueira, Alexandre Schwarzbald, Júlio

- Croda, Maurício Lacerda Nogueira, Gustavo Adolfo Sierra Romero, Quique Bassat, Cor Jesus Fontes, Bernardino Cláudio Albuquerque, Cláudio-Tadeu Daniel-Ribeiro, Wuelton Marcelo Monteiro, Marcus Vinícius Guimarães Lacerda, and for the CloroCovid-19 Team. Effect of High vs Low Doses of Chloroquine Diphosphate as Adjunctive Therapy for Patients Hospitalized With Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Infection: A Randomized Clinical Trial. *JAMA Network Open*, 3(4.23):e208857–e208857, 04 2020.
- [4] Christopher C Chang, Carson C Chow, Laurent Cam Tellier, Shashaank Vattikuti, Shaun M Purcell, and James J Lee. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1):1–16, 2015.
- [5] Zhaowei Chen, Jijia Hu, Zongwei Zhang, Shan Jiang, Shoumeng Han, Dandan Yan, Ruhong Zhuang, Ben Hu, and Zhan Zhang. Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial. *medRxiv*, 2020. doi:10.1101/2020.03.22.20040758.
- [6] P. Cingolani, A. Platts, I. e. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, 6(2):80–92, 2012.
- [7] T. G. Clark, A. E. Fry, S. Auburn, S. Campino, M. Diakite, A. Green, A. Richardson, Y. Y. Teo, K. Small, J. Wilson, M. Jallow, F. Sisay-Joof, M. Pinder, P. Sabeti, D. P. Kwiatkowski, and K. A. Rockett. Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. *Eur. J. Hum. Genet.*, 17(8):1080–1085, Aug 2009.
- [8] L. De Franceschi, E. Costa, F. Dima, M. Morandi, and O. Olivieri. Acute hemolysis by hydroxychloroquine was observed in G6PD-deficient patient with severe COVID-19 related lung injury. *Eur. J. Intern. Med.*, Apr 2020.
- [9] C. I. Fanello, C. Karema, P. Avellino, G. Bancone, A. Uwimana, S. J. Lee, U. d’Alessandro, and D. Modiano. High risk of severe anaemia after chlorproguanil-dapsone + artesunate

- antimalarial treatment in patients with G6PD (A-) deficiency. *PLoS ONE*, 3(12):e4031, 2008.
- [10] P. Gautret, J. C. Lagier, P. Parola, V. T. Hoang, L. Meddeb, M. Mailhe, B. Doudier, J. Courjon, V. Giordanengo, V. E. Vieira, H. T. Dupont, S. Honoré, P. Colson, E. Chabrière, B. La Scola, J. M. Rolain, P. Brouqui, and D. Raoult. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents*, page 105949, Mar 2020.
- [11] R. E. Howes, F. B. Piel, A. P. Patil, O. A. Nyangiri, P. W. Gething, M. Dewi, M. M. Hogg, K. E. Battle, C. D. Padilla, J. K. Baird, and S. I. Hay. G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Med.*, 9(11):e1001339, 2012.
- [12] S. Hwang, K. Mruk, S. Rahighi, A. G. Raub, C. H. Chen, L. E. Dorn, N. Horikoshi, S. Wakatsuki, J. K. Chen, and D. Mochly-Rosen. Correcting glucose-6-phosphate dehydrogenase deficiency with a small-molecule activator. *Nat Commun*, 9(1):4045, 10 2018.
- [13] Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D. Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallsabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeanet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, , Benjamin M.

- Neale, Mark J. Daly, and Daniel G. MacArthur. The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv*, 2020.
- [14] H. Li and R. Durbin. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5):589–595, Mar 2010.
- [15] X. Liu, C. Wu, C. Li, and E. Boerwinkle. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum. Mutat.*, 37(3):235–241, Mar 2016.
- [16] P. Loria, S. Miller, M. Foley, and L. Tilley. Inhibition of the peroxidative degradation of haem as the basis of action of chloroquine and other quinoline antimalarials. *Biochem. J.*, 339 (Pt 2):363–370, Apr 1999.
- [17] Joseph Magagnoli, Siddharth Narendran, Felipe Pereira, Tammy Cummings, James W Hardin, S Scott Sutton, and Jayakrishna Ambati. Outcomes of hydroxychloroquine usage in united states veterans hospitalized with covid-19. *medRxiv*, 2020.
- [18] S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N. Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N. Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F. Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C. Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S. Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M. S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E. Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G. Stamatoyannopoulos, J. T. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D. Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B. Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K. Thangaraj, S. P??bo, J. Kelso, N. Patterson, and D. Reich. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 538(7624):201–206, Oct 2016.

- [19] L. Manco, L. R. Botigué, M. L. Ribeiro, and A. Abade. G6PD deficient alleles and haplotype analysis of human G6PD locus in São Tomé and Príncipe (West Africa). *Hum. Biol.*, 79(6):679–686, Dec 2007.
- [20] A. Manjurano, N. Sepulveda, B. Nadjm, G. Mtove, H. Wangai, C. Maxwell, R. Olomi, H. Reyburn, E. M. Riley, C. J. Drakeley, and T. G. Clark. African glucose-6-phosphate dehydrogenase alleles associated with protection from severe malaria in heterozygous females in Tanzania. *PLoS Genet.*, 11(2):e1004960, Feb 2015.
- [21] P. J. Mason, J. M. Bautista, and F. Gilsanz. G6PD deficiency: the genotype-phenotype association. *Blood Rev.*, 21(5):267–283, Sep 2007.
- [22] E. C. Mbanefo, A. M. Ahmed, A. Titouna, A. Elmaraezy, N. T. Trang, N. Phuoc Long, N. Hoang Anh, T. Diem Nghi, B. The Hung, M. Van Hieu, N. Ky Anh, N. T. Huy, and K. Hirayama. Association of glucose-6-phosphate dehydrogenase deficiency and malaria: a systematic review and meta-analysis. *Sci Rep*, 7:45963, 04 2017.
- [23] E. M. McDonagh, C. F. Thorn, J. M. Bautista, I. Youngster, R. B. Altman, and T. E. Klein. PharmGKB summary: very important pharmacogene information for G6PD. *Pharmacogenet. Genomics*, 22(3):219–228, Mar 2012.
- [24] W. McLaren, L. Gil, S. E. Hunt, H. S. Riat, G. R. Ritchie, A. Thormann, P. Flicek, and F. Cunningham. The Ensembl Variant Effect Predictor. *Genome Biol.*, 17(1):122, 06 2016.
- [25] Mandeep R Mehra, Sapan S Desai, Frank Ruschitzka, and Amit N Patel. Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *The Lancet*, may 2020.
- [26] T. Motshoge, G. Ababio, L. Aleksenko, S. Souda, C. W. Muthoga, N. Mutukwa, L. Tawe, P. Ramatlho, L. Gabaitiri, S. Chihanga, T. Mosweunyane, S. Hamda, K. Moakofhi, D. Ntebela, E. Peloewetse, L. Mazhani, J. M. Pernica, J. Read, I. K. Quaye, and G. M. Paganotti. Prevalence of G6PD deficiency and associated haematological parameters in children from Botswana. *Infect. Genet. Evol.*, 63:73–78, 09 2018.

- [27] E.T. Nkhoma, C. Poole, V. Vannappagari, S.A. Hall, and E. Beutler. The global prevalence of glucose-6-phosphate dehydrogenase deficiency: a systematic review and meta-analysis. *Blood Cells, Molecules, and Diseases*, 41(3):267–78, 2009.
- [28] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [29] J. Raman, E. Allen, L. Workman, A. Mabuza, H. Swanepoel, G. Malatje, J. Freaan, L. Wiesner, and K. I. Barnes. Safety and tolerability of single low-dose primaquine in a low-intensity transmission area in South Africa: an open-label, randomized controlled trial. *Malar. J.*, 18(1):209, Jun 2019.
- [30] M. Ramsay, N. Crowther, E. Tambo, G. Agongo, V. Baloyi, S. Dikotope, X. Gómez-Olivé, N. Jaff, H. Sorgho, R. Wagner, C. Khayeka-Wandabwa, A. Choudhury, S. Hazelhurst, K. Kahn, Z. Lombard, F. Mukomana, C. Soo, H. Soodyall, A. Wade, S. Afolabi, I. Agorinya, L. Amenga-Etego, S. A. Ali, J. D. Bognini, R. P. Boua, C. Debpuur, S. Diallo, E. Fato, A. Kazienga, S. Z. Konkobo, P. M. Kouraogo, F. Mashinya, L. Micklesfield, S. Nakanabo-Diallo, B. Njamwea, E. Nonterah, S. Ouedraogo, V. Pillay, A. M. Somande, P. Tindana, R. Twine, M. Alberts, C. Kyobutungi, S. A. Norris, A. R. Oduro, H. Tinto, S. Tollman, and O. Sankoh. H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Global Health, Epidemiology and Genomics*, 1:e20, 2016.
- [31] Carlos HM Rodrigues, Douglas EV Pires, and David B Ascher. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic Acids Research*, 46(W1):W350–W355, 04 2018.
- [32] J. M. Sanders, M. L. Monogue, T. Z. Jodlowski, and J. B. Cutrell. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*, Apr 2020.
- [33] M. A. Saunders, M. Slatkin, C. Garner, M. F. Hammer, and M. W. Nachman. The extent of linkage disequilibrium caused by selection on G6PD in humans. *Genetics*, 171(3):1219–1229, Nov 2005.

- [34] S. A. Shekalaghe, R. ter Braak, M. Daou, R. Kavishe, W. van den Bijllaardt, S. van den Bosch, J. B. Koenderink, A. J. Luty, C. J. Whitty, C. Drakeley, R. W. Sauerwein, and T. Bousema. In Tanzania, hemolysis after a single dose of primaquine coadministered with an artemisinin is not restricted to glucose-6-phosphate dehydrogenase-deficient (G6PD A-) individuals. *Antimicrob. Agents Chemother.*, 54(5):1762–1768, May 2010.
- [35] F. S. Taccone, J. Gorham, and J. L. Vincent. Hydroxychloroquine in the management of critically ill patients with COVID-19: the need for an evidence base. *Lancet Respir Med*, Apr 2020.
- [36] US Food and Drug Administration. US Food and Drug Administration Chloroquine Label. https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/006002s0441b1.pdf, March 2017. Online, accessed 10 April.
- [37] US Food and Drug Administration. US Food and Drug Administration Hydroxychloroquine Label. https://www.accessdata.fda.gov/drugsatfda_docs/label/2007/009768s0411b1.pdf, March 2017. Online, accessed 10 April.
- [38] T. J. Vulliamy, M. D’Urso, G. Battistuzzi, M. Estrada, N. S. Foulkes, G. Martini, V. Calabro, V. Poggi, R. Giordano, and M. Town. Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. *Proc. Natl. Acad. Sci. U.S.A.*, 85(14):5171–5175, Jul 1988.
- [39] T. J. Vulliamy, A. Othman, M. Town, A. Nathwani, A. G. Falusi, P. J. Mason, and L. Luzzatto. Polymorphic sites in the African population detected by sequence analysis of the glucose-6-phosphate dehydrogenase gene outline the evolution of the variants A and A-. *Proc. Natl. Acad. Sci. U.S.A.*, 88(19):8568–8571, Oct 1991.
- [40] J. Wang, Q. Z. Xiao, Y. M. Chen, S. Yi, D. Liu, Y. H. Liu, C. M. Zhang, X. F. Wei, Y. Q. Zhou, X. M. Zhong, C. Y. Zhao, F. Xiong, X. C. Wei, and X. M. Xu. DNA hypermethylation and X chromosome inactivation are major determinants of phenotypic variation in women heterozygous for G6PD mutations. *Blood Cells Mol. Dis.*, 53(4):241–245, Dec 2014.

- [41] Yi-Hsuan Wu, Ching-Ping Tseng, Mei-Ling Cheng, Hung-Yao Ho, Shin-Ru Shih, and Daniel Tsun-Yee Chiu. Glucose-6-Phosphate Dehydrogenase Deficiency Enhances Human Coronavirus 229E Infection. *The Journal of Infectious Diseases*, 197(6):812–816, 03 2008.
- [42] F. Yu, S. Zhang, B. Chen, Y. Zhou, C. Ma, S. Yang, Y. Tang, D. Huang, X. Xie, Q. Xiao, and L. Wang. Evaluation of the Diagnostic Accuracy of the CareStart Glucose-6-Phosphate Dehydrogenase Deficiency Rapid Diagnostic Test among Chinese Newborns. *J. Trop. Pediatr.*, Feb 2020.
- [43] Y. Zhou, S. Mkrтчian, M. Kumondai, M. Hiratsuka, and V. M. Lauschke. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.*, 19(2):115–126, 04 2019.